*Proc. R. Soc. Lond.* A **392**, 45–57 (1984)
*Printed in Great Britain*

# Quantal phase factors accompanying adiabatic changes

BY M. V. BERRY, F.R.S.
*H. H. Wills Physics Laboratory, University of Bristol, Tyndall Avenue, Bristol BS8 1TL, U.K.*

(*Received* 13 *June* 1983)

A quantal system in an eigenstate, slowly transported round a circuit C by varying parameters $\boldsymbol{R}$ in its Hamiltonian $\hat{H}(\boldsymbol{R})$, will acquire a geometrical phase factor $\exp\{i\gamma(C)\}$ in addition to the familiar dynamical phase factor. An explicit general formula for $\gamma(C)$ is derived in terms of the spectrum and eigenstates of $\hat{H}(\boldsymbol{R})$ over a surface spanning C. If C lies near a degeneracy of $\hat{H}$, $\gamma(C)$ takes a simple form which includes as a special case the sign change of eigenfunctions of real symmetric matrices round a degeneracy. As an illustration $\gamma(C)$ is calculated for spinning particles in slowly-changing magnetic fields; although the sign reversal of spinors on rotation is a special case, the effect is predicted to occur for bosons as well as fermions, and a method for observing it is proposed. It is shown that the Aharonov–Bohm effect can be interpreted as a geometrical phase factor.

## 1. INTRODUCTION

Imagine a quantal system whose Hamiltonian $\hat{H}$ describes the effects of an unchanging environment, and let the system be in a stationary state. If the environment, and hence $\hat{H}$, is slowly altered, it follows from the adiabatic theorem (Messiah 1962) that at any instant the system will be in an eigenstate of the instantaneous $\hat{H}$. In particular, if the Hamiltonian is returned to its original form the system will return to its original state, apart from a phase factor. This phase factor is observable by interference if the cycled system is recombined with another that was separated from it at an earlier time and whose Hamiltonian was kept constant.

My purpose here is to explain how the phase factor contains a circuit-dependent component $\exp(i\gamma)$ in addition to the familiar dynamical component $\exp(-iEt/\hbar)$ which accompanies the evolution of any stationary state. A general formula for $\gamma$ in terms of the eigenstates of $\hat{H}$ will be obtained in § 2. If the circuit is close to a degeneracy in the spectrum of $\hat{H}$, $\gamma$ takes a particularly simple form which will be derived in § 3; this contains, as a special case, the sign change around a degeneracy of the eigenstates of a system whose Hamiltonian is real as well as Hermitian (Herzberg & Longuet-Higgins 1963; Longuet-Higgins 1975; Mead 1979; Mead & Truhlar 1979; Mead 1980*a*, *b*; Berry & Wilkinson 1984).

A particle of any spin in an eigenstate of a slowly-rotated magnetic field is another case where $\gamma$ can be calculated explicitly (§ 4), and gives predictions that could be

tested experimentally. This phase factor exists for bosons as well as fermions. A special case is the sign change of spinors slowly rotated by $2\pi$, predicted by Aharonov & Susskind (1967); this will be shown to be different from the dynamical phase factors measured in experiments on precessing neutrons (reviewed by Silverman 1980).

Finally, it is shown in § 5 that physical effects of magnetic vector potentials in the absence of fields, predicted by Aharonov & Bohm (1959) and observed by Chambers (1960), can be understood as special cases of the geometrical phase factor.

## 2. General formula for phase factor

Let the Hamiltonian $\hat{H}$ be changed by varying parameters $\boldsymbol{R} = (X, Y, \ldots)$ on which it depends. Then the excursion of the system between times $t = 0$ and $t = T$ can be pictured as transport round a closed path $\boldsymbol{R}(t)$ in parameter space, with Hamiltonian $\hat{H}(\boldsymbol{R}(t))$ and such that $\boldsymbol{R}(T) = \boldsymbol{R}(0)$. The path will henceforth be called a circuit and denoted by C. For the adiabatic approximation to apply, $T$ must be large.

The state $|\psi(t)\rangle$ of the system evolves according to Schrödinger's equation

$$\hat{H}(\boldsymbol{R}(t))\,|\psi(t)\rangle = \mathrm{i}\hbar\,|\dot{\psi}(t)\rangle. \tag{1}$$

At any instant, the natural basis consists of the eigenstates $|n(\boldsymbol{R})\rangle$ (assumed discrete) of $\hat{H}(\boldsymbol{R})$ for $\boldsymbol{R} = \boldsymbol{R}(t)$, that satisfy

$$\hat{H}(\boldsymbol{R})\,|n(\boldsymbol{R})\rangle = E_n(\boldsymbol{R})\,|n(\boldsymbol{R})\rangle, \tag{2}$$

with energies $E_n(\boldsymbol{R})$. This eigenvalue equation implies no relation between the phases of the eigenstates $|n(\boldsymbol{R})\rangle$ at different $\boldsymbol{R}$. For present purposes any (differentiable) choice of phases can be made, provided $|n(\boldsymbol{R})\rangle$ is single-valued in a parameter domain that includes the circuit C.

Adiabatically, a system prepared in one of these states $|n(\boldsymbol{R}(0))\rangle$ will evolve with $\hat{H}$ and so be in the state $|n(\boldsymbol{R}(t))\rangle$ at $t$.

Thus $|\psi\rangle$ can be written as

$$|\psi(t)\rangle = \exp\left\{\frac{-\mathrm{i}}{\hbar}\int_0^t \mathrm{d}t' E_n(\boldsymbol{R}(t'))\right\}\exp\left(\mathrm{i}\gamma_n(t)\right)|n(\boldsymbol{R}(t))\rangle. \tag{3}$$

The first exponential is the familiar dynamical phase factor. In this paper the object of attention is the second exponential. The crucial point will be that its phase $\gamma_n(t)$ is *non-integrable*; $\gamma_n$ cannot be written as a function of $\boldsymbol{R}$ and in particular is not single-valued under continuation around a circuit, i.e. $\gamma_n(T) \neq \gamma_n(0)$.

The function $\gamma_n(t)$ is determined by the requirement that $|\psi(t)\rangle$ satisfy Schrödinger's equation, and direct substitution of (3) into (1) leads to

$$\dot{\gamma}_n(t) = \mathrm{i}\langle n(\boldsymbol{R}(t))\,|\,\nabla_{\boldsymbol{R}} n(\boldsymbol{R}(t))\rangle\cdot\dot{\boldsymbol{R}}(t). \tag{4}$$

The total phase change of $|\psi\rangle$ round C is given by

$$|\psi(T)\rangle = \exp(\mathrm{i}\gamma_n(\mathrm{C}))\exp\left\{\frac{-\mathrm{i}}{\hbar}\int_0^T \mathrm{d}t\, E_n(\boldsymbol{R}(t))\right\}|\psi(0)\rangle, \tag{5}$$

where the *geometrical phase change* is

$$\gamma_n(\mathrm{C}) = \mathrm{i}\oint_{\mathrm{C}} \langle n(\boldsymbol{R})|\,\nabla_{\boldsymbol{R}}\, n(\boldsymbol{R})\rangle\cdot \mathrm{d}\boldsymbol{R}. \tag{6}$$

Thus $\gamma_n(\mathrm{C})$ is given by a circuit integral in parameter space and is independent of how the circuit is traversed (provided of course that this is slow enough for the adiabatic approximation to hold). The normalization of $|n\rangle$ implies that $\langle n|\nabla_{\boldsymbol{R}} n\rangle$ is imaginary, which guarantees that $\gamma_n$ is real.

Direct evaluation of $|\nabla_{\boldsymbol{R}} n\rangle$ requires a locally single-valued basis for $|n\rangle$ and can be awkward. Such difficulties are avoided by transforming the circuit integral (6) into a surface integral over any surface in parameter space whose boundary is C. In order to employ familiar vector calculus, parameter space will be considered as three-dimensional, and this will turn out to be the important case in applications; the generalization to higher dimensions will be outlined at the end of this section.

Stokes's theorem applied to (6) gives, in an obvious abbreviated notation.

$$\gamma_n(\mathrm{C}) = -\mathrm{Im}\iint_{\mathrm{C}} \mathrm{d}\boldsymbol{S}\cdot\nabla\times\langle n|\,\nabla n\rangle, \tag{7a}$$

$$= -\mathrm{Im}\iint_{\mathrm{C}} \mathrm{d}\boldsymbol{S}\cdot\langle\nabla n|\times|\nabla n\rangle, \tag{7b}$$

$$= -\mathrm{Im}\iint_{\mathrm{C}} \mathrm{d}\boldsymbol{S}\cdot\sum_{m\neq n}\langle\nabla n|\,m\rangle\times\langle m\,|\nabla n\rangle, \tag{7c}$$

where $\mathrm{d}\boldsymbol{S}$ denotes area element in $\boldsymbol{R}$ space and the exclusion in the summation is justified by $\langle n|\,\nabla n\rangle$ being imaginary. The off-diagonal elements are obtained from (2) as

$$\langle m|\,\nabla n\rangle = \langle m|\,\nabla\hat{H}\,|n\rangle/(E_n - E_m), \quad m\neq n. \tag{8}$$

Thus $\gamma_n$ can be expressed as

$$\gamma_n(\mathrm{C}) = -\iint_{\mathrm{C}} \mathrm{d}\boldsymbol{S}\cdot\boldsymbol{V}_n(\boldsymbol{R}), \tag{9}$$

where

$$\boldsymbol{V}_n(\boldsymbol{R}) \equiv \mathrm{Im}\sum_{m\neq n}\frac{\langle n(\boldsymbol{R})|\,\nabla_{\boldsymbol{R}}\hat{H}(\boldsymbol{R})|\,m(\boldsymbol{R})\rangle\times\langle m(\boldsymbol{R})|\,\nabla_{\boldsymbol{R}}\hat{H}(\boldsymbol{R})|\,n(\boldsymbol{R})\rangle}{(E_m(\boldsymbol{R})-E_n(\boldsymbol{R}))^2}. \tag{10}$$

Obviously $\gamma_n(\mathrm{C})$ is zero for a circuit which retraces itself and so encloses no area.

Equations (9) and (10) embody the central results of this paper. Because the dependence on $|\nabla n\rangle$ has been eliminated, phase relations between eigenstates with different parameters are now immaterial, and (as is evident from the form of (10)), it is no longer necessary to choose $|m\rangle$ and $|n\rangle$ to be single-valued in $\boldsymbol{R}$: any solutions of (2) may be employed without affecting the value of $\boldsymbol{V}_n$. This is a surprising conclusion, as can be seen by comparing (9) with (7*a*) which show that $\boldsymbol{V}_n$ is the curl of a vector, $\langle n|\nabla n\rangle$, and $\langle n|\nabla n\rangle$ certainly does depend on the choice of phase

of the (single-valued) eigenstate $|n(\boldsymbol{R})\rangle$. The dependence on phase is of the following kind: if $|n\rangle \to \exp\{\mathrm{i}\mu(\boldsymbol{R})\}|n\rangle$ then $\langle n|\nabla n\rangle \to \langle n|\nabla n\rangle + \mathrm{i}\nabla\mu$ (in another context the importance of such gauge transformations has been emphasized by Wu & Yang (1975)). Thus the vector is not unique but its curl is. The quantity $\boldsymbol{V}_n$ is analogous to a 'magnetic field' (in parameter space) whose 'vector potential' is $\mathrm{Im}\langle n|\nabla n\rangle$. In Appendix A it is shown directly from (10) that $\nabla\cdot\boldsymbol{V}_n$ vanishes, thus confirming that (9) gives a unique value for $\gamma_n(\mathrm{C})$.

Using perturbation theory, Mead & Truhlar (1979) obtained essentially the formulae (9) and (10) for an infinitesimal circuit, in a study of molecular electronic states which (in the Born–Oppenheimer approximation) depend parametrically on nuclear coordinates. Their phase factor was not intended to apply to a $|\psi\rangle$ that evolves slowly under the time-dependent Schrödinger equation, but to the variation of eigenstates $|n\rangle$ under a particular phase-continuation rule in $\boldsymbol{R}$-space which can be shown to give the same result.

In parameter spaces of higher dimension, Stokes's theorem cannot be employed to transform the circuit integral (6). The appropriate generalization, provided by the theory of differential forms (see, for example, Arnold 1978, chap. 7), transforms (6) into the integral of a 2-form over a surface bounded by C. The surprising result (10) can now be expressed as follows: independently of the choice of phases of the eigenstates, there exists in parameter space a *phase 2-form*, which gives $\gamma(\mathrm{C})$ when integrated over any surface spanning C. This 2-form is obtained from (10) by replacing $\nabla$ by the exterior derivative d and $\times$ by the wedge product $\wedge$. The validity of this generalization is consistent with the observation that in the three-dimensional version there are infinitely many choices of interpolating Hamiltonian (and hence of parameter spaces) on the surfaces bounded by C, and the geometrical phase factor is independent of the choice.

Professor Barry Simon (1983), commenting on the original version of this paper, points out that the geometrical phase factor has a mathematical interpretation in terms of holonomy, with the phase two-form emerging naturally (in the form (7*b*)) as the curvature (first Chern class) of a Hermitian line bundle.

## 3. Degeneracies

The energy denominators in (10) show that if the circuit C lies close to a point $\boldsymbol{R}^*$ in parameter space at which the state $n$ is involved in a degeneracy, then $\boldsymbol{V}_n(\boldsymbol{R})$, and hence $\gamma_n(\mathrm{C})$, is dominated by the terms $m$ corresponding to the other states involved. We shall consider the commonest situation, where the degeneracy involves only two states, to be denoted $+$ and $-$, where $E_+(\boldsymbol{R}) \geqslant E_-(\boldsymbol{R})$. For $\boldsymbol{R}$ near $\boldsymbol{R}^*$, $\hat{H}(\boldsymbol{R})$ can be expanded to first order in $\boldsymbol{R}-\boldsymbol{R}^*$, and

$$\boldsymbol{V}_+(\boldsymbol{R}) = \mathrm{Im}\,\frac{\langle +(\boldsymbol{R})|\,\nabla\hat{H}(\boldsymbol{R}^*)|-(\boldsymbol{R})\rangle \times \langle -(\boldsymbol{R})|\,\nabla\hat{H}(\boldsymbol{R}^*)|+(\boldsymbol{R})\rangle}{(E_+(\boldsymbol{R})-E_-(\boldsymbol{R}))^2}. \tag{11}$$

Obviously $\boldsymbol{V}_-(\boldsymbol{R}) = -\boldsymbol{V}_+(\boldsymbol{R})$, so that $\gamma_-(\mathrm{C}) = -\gamma_+(\mathrm{C})$.

Without essential loss of generality we can take $E_{\pm}(\boldsymbol{R}^*) = 0$ and $\boldsymbol{R}^* = 0$. $H(\boldsymbol{R})$ can be represented by a $2 \times 2$ Hermitian matrix coupling the two states. The most general such matrix satisfying the given conditions depends on three parameters $X$, $Y$, $Z$ which will be taken as components of $\boldsymbol{R}$, and by linear transformation in $\boldsymbol{R}$-space can be brought into the following standard form

$$\hat{H}(\boldsymbol{R}) = \frac{1}{2}\begin{bmatrix} Z & X - \mathrm{i}Y \\ X + \mathrm{i}Y & -Z \end{bmatrix}. \tag{12}$$

The eigenvalues are

$$E_+(\boldsymbol{R}) = -E_-(\boldsymbol{R}) = \tfrac{1}{2}(X^2 + Y^2 + Z^2)^{\frac{1}{2}} = \tfrac{1}{2}R. \tag{13}$$

Thus the degeneracy is an isolated point at which all three parameters vanish. This illustrates an old result of Von Neumann & Wigner (1929): for generic Hamiltonians (Hermitian matrices), it is necessary to vary three parameters in order to make a degeneracy occur accidentally, that is, not on account of symmetry. Alternatively stated, degeneracies have co-dimension three.

The form (12) was chosen to exploit the fact that

$$\nabla \hat{H} = \tfrac{1}{2}\hat{\boldsymbol{\sigma}}, \tag{14}$$

where the components $\hat{\sigma}_X$, $\hat{\sigma}_Y$, $\hat{\sigma}_Z$ of the vector operator $\hat{\boldsymbol{\sigma}}$ are the Pauli spin matrices. When evaluating the matrix elements in (11) it greatly simplifies the calculations to take advantage of the isotropy of spin and temporarily rotate axes so that the $Z$-axis points along $\boldsymbol{R}$, and to employ the following relations, which come from the commutation laws between the components of $\hat{\boldsymbol{\sigma}}$:

$$\hat{\sigma}_X |\pm\rangle = |\mp\rangle, \quad \hat{\sigma}_Y |\pm\rangle = \pm \mathrm{i} |\mp\rangle, \quad \hat{\sigma}_Z |\pm\rangle = \pm |\pm\rangle. \tag{15}$$

With these rotated axes, (11) gives

$$\left.\begin{aligned} V_{X+} &= (\mathrm{Im}\langle + | \hat{\sigma}_Y | - \rangle \langle - | \hat{\sigma}_Z | + \rangle)/2R^2 = 0, \\ V_{Y+} &= (\mathrm{Im}\langle + | \hat{\sigma}_Z | - \rangle \langle - | \hat{\sigma}_X | + \rangle)/2R^2 = 0, \\ V_{Z+} &= \mathrm{Im}\langle + | \hat{\sigma}_X | - \rangle \langle - | \hat{\sigma}_Y | + \rangle = 1/2R^2. \end{aligned}\right\} \tag{16}$$

Reverting to unrotated axes, we obtain

$$\boldsymbol{V}_+(\boldsymbol{R}) = \boldsymbol{R}/2R^3. \tag{17}$$

Now use of (9) shows that the phase change $\gamma_+(\mathrm{C})$ is the flux through C of the magnetic field of a monopole with strength $-\frac{1}{2}$ located at the degeneracy. Thus we obtain the pleasant result, valid for the natural choice (12) of standard form for $\hat{H}$, that the geometrical phase factor associated with C is

$$\exp\{\mathrm{i}\gamma_\pm(\mathrm{C})\} = \exp\{\mp \tfrac{1}{2}\mathrm{i}\Omega(\mathrm{C})\}, \tag{18}$$

where $\Omega(\mathrm{C})$ is the *solid angle* that C subtends at the degeneracy; $\Omega$ is, in a sense, a measure of the *view* of the circuit as seen from the degeneracy. The phase factor is

independent of the choice of surface spanning C, because $\Omega$ can change only in multiples of $4\pi$ (when the surface is deformed to pass through the degeneracy).

An important special case of (18) occurs when C consists entirely of *real* Hamiltonians and so is confined to the plane $Y = 0$ (cf. (12)). The energy levels $E_\pm$ intersect conically in the space $E$, $X$, $Z$, whose origin, where the degeneracy occurs, is a 'diabolical point' of the type recently studied by Berry & Wilkinson (1984) in the spectra of triangles. This illustrates the result that for real symmetric matrices, degeneracies have co-dimension two: see Appendix 10 of Arnold 1978. If C encloses the degeneracy, $\Omega = \pm 2\pi$; if not, $\Omega = 0$. Thus the phase factor (18) is

$$\begin{aligned} \exp\{\mathrm{i}\gamma_\pm(\mathrm{C})\} &= -1, \quad \text{if C encircles the degeneracy,} \\ &= +1, \quad \text{otherwise,} \end{aligned} \tag{19}$$

which expresses the sign changes of real wavefunctions as a degeneracy involving them is encircled, a phenomenon first described by Herzberg & Longuet-Higgins (1963). (Sign changes are not restricted to circuits involving real Hamiltonians: (18) shows that the phase factor is $-1$ if C lies in *any* plane through the degeneracy and encircles it.)

Confirmation of the correctness of (17) can be obtained without the rotation-of-axes trick, by a lengthy calculation of (11) involving explicit formulae for the eigenvectors $|\pm(\boldsymbol{R})\rangle$ of the matrix (12). Alternatively, direct continuation of the eigenvectors may be attempted. This cannot be accomplished for all circuits by means of (6) because it is not possible to construct eigenvectors that are globally single-valued continuous functions of $\boldsymbol{R}$; multivaluedness can be reduced to singular lines connecting the degeneracy with infinity, and in the analogue $\boldsymbol{V}(\boldsymbol{R})$ these appear as Dirac strings attached to the monopole. Such approaches obscure the simplicity and essential isotropy of the solid-angle result (17).

Using topological arguments not involving explicit formulae for $\gamma_n(\mathrm{C})$, Stone (1976) proved that if C is expanded from one point $\boldsymbol{R}$ and contracted on to another so as to sweep out a surface enclosing a degeneracy, then the geometrical phase factor traverses a circle in its Argand plane. This property (which follows easily from (18)), is the Hermitian generalization of the sign-reversal test for degeneracy.

## 4. Spins in magnetic fields

A particle with spin $s$ (integer or half-integer) interacts with a magnetic field $\boldsymbol{B}$ via the Hamiltonian

$$\hat{H}(\boldsymbol{B}) = \kappa\hbar\boldsymbol{B}\cdot\hat{\boldsymbol{s}}, \tag{20}$$

where $\kappa$ is a constant involving the gyromagnetic ratio and $\hat{\boldsymbol{s}}$ is the vector spin operator with $2s+1$ eigenvalues $n$ with integer spacing and that lie between $-s$ and $+s$. The eigenvalues are

$$E_n(\boldsymbol{B}) = \kappa\hbar Bn, \tag{21}$$

and so there is a $(2s+1)$-fold degeneracy when $\boldsymbol{B} = 0$. (The special case $s = \frac{1}{2}$ reproduces the two-fold degeneracy considered in the last section.) We consider

the components of $\boldsymbol{B}$ as the parameters $\boldsymbol{R}$ in our previous analysis, and calculate the phase change $\gamma_n(\mathrm{C})$ of an eigenstate $|n, s(\boldsymbol{B})\rangle$ of $\hat{\boldsymbol{s}}$ in the direction along $\boldsymbol{B}$, as $\boldsymbol{B}$ is slowly varied (and hence the spin rotated) round a circuit C.

The vector $\boldsymbol{V}_n(\boldsymbol{B})$ as given by (10) can be expressed by using (20) and (21) as

$$\boldsymbol{V}_n(B) = \frac{\mathrm{Im}}{B^2} \sum_{m \neq n} \frac{\langle n, s(\boldsymbol{B}) | \hat{\boldsymbol{s}} | m, s(\boldsymbol{B})\rangle \times \langle m, s(\boldsymbol{B}) | \hat{\boldsymbol{s}} | n, s(\boldsymbol{B})\rangle}{(m-n)^2}. \tag{22}$$

To evaluate the matrix elements we again temporarily rotate axes so that the $Z$-axis points along $\boldsymbol{B}$, and employ the following generalization of (15):

$$\left.\begin{aligned} (\hat{s}_X + \mathrm{i}\hat{s}_Y)\,|n, s\rangle &= [s(s+1) - n(n+1)]^{\frac{1}{2}}\,|n+1, s\rangle, \\ (\hat{s}_X - \mathrm{i}\hat{s}_Y)\,|n, s\rangle &= [s(s+1) - n(n-1)]^{\frac{1}{2}}\,|n-1, s\rangle, \\ s_Z\,|n, s\rangle &= n\,|n, s\rangle. \end{aligned}\right\} \tag{23}$$

It is clear that only states with $m = n \pm 1$ are coupled with $|n\rangle$ in (22), and that $V_x$ and $V_y$ are zero because they involve off-diagonal elements of $\hat{s}_Z$. To find $V_Z$, we make use of (23) to obtain

$$\left.\begin{aligned} \langle n \pm 1, s|\, s_X\,|n, s\rangle &= \tfrac{1}{2}[s(s+1) - n(n \pm 1)]^{\frac{1}{2}}, \\ \langle n \pm 1, s|\, s_Y\,|n, s\rangle &= \mp \tfrac{1}{2}\mathrm{i}[s(s+1) - n(n \pm 1)]^{\frac{1}{2}}, \end{aligned}\right\} \tag{24}$$

then (22) gives

$$\begin{aligned} \boldsymbol{V}_{Zn} &= \frac{\mathrm{Im}}{B^2}\{\langle n|\, s_X\,|n+1\rangle\langle n+1|\, s_Y\,|n\rangle - \langle n|\, s_Y\,|n+1\rangle\langle n+1|\, s_X\,|n\rangle \\ &\quad + \langle n|\, s_X\,|n-1\rangle\langle n-1|\, s_Y\,|n\rangle - \langle n|\, s_Y\,|n-1|\rangle\langle n-1|\, s_X\,|n\rangle\} \\ &= \frac{n}{B^2}. \end{aligned} \tag{25}$$

Reverting to unrotated axes, we obtain

$$\boldsymbol{V}_n(\boldsymbol{B}) = n\boldsymbol{B}/B^3. \tag{26}$$

Now, use of (9) shows that $\gamma_n(\mathrm{C})$ is the flux through C of the 'magnetic field' of a monopole $-n$ located at the origin of magnetic field space. Thus the geometrical phase factor is

$$\exp\{\mathrm{i}\gamma_n(\mathrm{C})\} = \exp\{-\mathrm{i}n\Omega(\mathrm{C})\}, \tag{27}$$

where $\Omega(\mathrm{C})$ is the solid angle that C subtends at $\boldsymbol{B} = 0$. Note that $\gamma_n$ depends only on the eigenvalue $n$ of the spin component along $\boldsymbol{B}$ and not on the spin $s$ of the particle, so that $\gamma_n$ is insensitive to the strength $2s+1$ of the degeneracy at $\boldsymbol{B} = 0$.

It follows from (27) that any phase change can be produced by varying $\boldsymbol{B}$ round a suitable circuit. For fermions (half-integer $n$), a whole turn of $\boldsymbol{B}$ (rotation through $2\pi$ in a plane, giving $\Omega = 2\pi$) produces a phase factor $-1$. In the special case $n = \frac{1}{2}$ this shows that the sign change of spinors on rotation and the sign change of wavefunctions round a degeneracy have the same mathematical origin. For bosons (integer $n$), a whole turn of $\boldsymbol{B}$ produces a phase factor $+1$. To produce a sign change,

different circuits are required; if $n = 1$, for example, varying $\boldsymbol{B}$ round a cone of semiangle 60° will give $\Omega = \gamma = \pi$ and hence a phase factor $-1$.

The following experiment could be carried out to test the predictions embodied in (27). A polarized monoenergetic beam of particles in spin state $n$ along a magnetic field $\boldsymbol{B}$ is split into two. Along the path of one beam $\boldsymbol{B}$ is kept constant. Along the path of the other beam, $\boldsymbol{B}$ is kept constant in magnitude but its direction is varied slowly (in comparison with the dynamical precession frequency) round a circuit C subtending a solid angle $\Omega$, the field being generated by an arrangement enabling $\Omega$ to be changed. The beams are then combined and the count rate at a detector is measured as a function of $\Omega$. The dynamical phase factor (the second exponential in (5) is the same for both beams because the energy $E_n(\boldsymbol{B})$ (21) is insensitive to the direction of $\boldsymbol{B}$. There will in addition be a propagation phase factor which can be made unity by adjusting the path-length of one of the beams when $\Omega = 0$. The resulting fringes occur as a consequence of the geometrical phase factor. If C is a circuit round a cone of semiangle $\theta$, the predicted intensity contrast is

$$I(\theta) = \cos^2(n\pi(1-\cos\theta)). \tag{28}$$

I wish to emphasize that this proposed experiment is different from those carried out by Rauch *et al.* (1975, 1978) and Werner *et al.* (1975) (see Silverman 1980) with *unpolarized* neutrons in a *constant* magnetic field. Those neutrons were not in an eigenstate, and their phase changed dynamically, rather than geometrically, under the Hamiltonian (20) (with $\boldsymbol{B}$ along $Z$ and $\hat{\boldsymbol{\sigma}}$ replacing $\hat{\boldsymbol{s}}$) according to the evolution operator

$$\exp(-\mathrm{i}\hat{H}t/\hbar) = \exp(-B\kappa t\hat{\sigma}_Z) = \cos\tfrac{1}{2}\kappa Bt\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix} + \mathrm{i}\sin\tfrac{1}{2}\kappa Bt\begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix}. \tag{29}$$

The sign changed whenever $\frac{1}{2}\kappa Bt$ was an odd multiple of $\pi$, and this was interpreted on the basis of precession theory as corresponding to odd numbers of complete rotations about $\boldsymbol{B}$.

## 5. Aharonov–Bohm effect

Consider a magnetic field consisting of a single line with flux $\Phi$. For positions $\boldsymbol{R}$ not on the flux line, the field is zero but there must be a vector potential $\boldsymbol{A}(\boldsymbol{R})$ satisfying

$$\oint_{\mathrm{C}} \boldsymbol{A}(\boldsymbol{R})\cdot\mathrm{d}\boldsymbol{R} = \Phi, \tag{30}$$

for circuits C threaded by the flux line. Aharonov & Bohm (1959) showed that in quantum mechanics such vector potentials have physical significance even though they correspond to zero field. I shall now show how their effect can be interpreted as a geometrical phase change of the type described in § 2.

Let the quantal system consist of particles with charge $q$ confined to a box situated at $\boldsymbol{R}$ and not penetrated by the flux line (figure 1). In the absence of flux

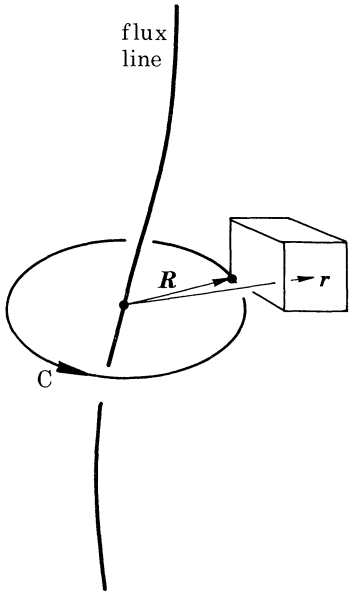


FIGURE 1. Aharonov–Bohm effect in a box transported round a flux line.

($\boldsymbol{A} = 0$), the particle Hamiltonian depends on position $\hat{\boldsymbol{r}}$ and conjugate momentum $\hat{\boldsymbol{p}}$ as follows:

$$\hat{H} = H(\hat{\boldsymbol{p}}, \hat{\boldsymbol{r}} - \boldsymbol{R}), \tag{31}$$

and the wavefunctions have the form $\psi_n(\boldsymbol{r} - \boldsymbol{R})$ with energies $E_n$ independent of $\boldsymbol{R}$. With non-zero flux, the states $|n(\boldsymbol{R})\rangle$ satisfy

$$H(\hat{\boldsymbol{p}} - q\boldsymbol{A}(\hat{\boldsymbol{r}}), \hat{\boldsymbol{r}} - \boldsymbol{R})|\, n(\boldsymbol{R})\rangle = E_n\, |n(\boldsymbol{R})\rangle, \tag{32}$$

an equation whose exact solutions are obtained by multiplying $\psi_n$ by an appropriate Dirac phase factor, giving

$$\langle \boldsymbol{r}|\, n(\boldsymbol{R})\rangle = \exp\left\{\frac{\mathrm{i}q}{\hbar}\int_{\boldsymbol{R}}^{\boldsymbol{r}} \mathrm{d}\boldsymbol{r}' \cdot \boldsymbol{A}(\boldsymbol{r}')\right\} \psi_n(\boldsymbol{r} - \boldsymbol{R}). \tag{33}$$

These solutions are single-valued in $\boldsymbol{r}$ and (locally) in $\boldsymbol{R}$. The energies are unaffected the vector potential.

Now let the box be transported round a circuit C threaded by the flux line; in this particular case it is not necessary that the transport be adiabatic. After completion of the circuit there will be a geometrical phase change that can be calculated from (6) and (33) by using

$$\begin{aligned}\langle n(\boldsymbol{R})|\, \nabla_{\boldsymbol{R}} n(\boldsymbol{R})\rangle &= \iiint \mathrm{d}^3\boldsymbol{r}\psi_n^*(\boldsymbol{r} - \boldsymbol{R})\left\{\frac{-\mathrm{i}q}{\hbar} A(\boldsymbol{R})\,\psi_n(\boldsymbol{r} - \boldsymbol{R}) + \nabla_{\boldsymbol{R}}\psi_n(\boldsymbol{r} - \boldsymbol{R})\right\} \\ &= -\mathrm{i}q\boldsymbol{A}(\boldsymbol{R})/\hbar. \end{aligned} \tag{34}$$

(The vanishing of the second term in braces follows from the normalization of $\psi_n$.) Evidently in this example the analogy between $\mathrm{Im}\langle n|\nabla n\rangle$ and a magnetic vector potential becomes a reality. Thus

$$\gamma_n(\mathrm{C}) = \frac{q}{\hbar}\oint_{\mathrm{C}} \boldsymbol{A}(\boldsymbol{R})\cdot \mathrm{d}\boldsymbol{R} = q\Phi/\hbar, \tag{35}$$

which shows that the phase factor is independent of $n$, and also of C if this winds once round the flux line. The phase factor could be observed by interference between the particles in the transported box and those in a box which was not transported round the circuit.

In elementary presentations of the Aharonov–Bohm effect (including its anticipation by Ehrenburg & Siday 1949), the Dirac phase factor is often invoked in comparing systems passing opposite sides of a flux line. Such invocations are subject to the objection that the wavefunction thus obtained is not single-valued. One way to avoid this objection is by summation over all contributions (whirling waves) representing different windings round the flux line (Schulman 1981; Berry 1980; Morandi & Menossi 1984). Another way, adopted in the original paper by Aharonov & Bohm, is to solve Schrödinger's equation exactly for scattering in the flux line's vector potential. The argument of the preceding paragraphs, which employs the geometrical phase factor, is a third way of obtaining the Aharonov–Bohm effect by using only single-valued wavefunctions.

Mead (1980*a*, *b*), employs the term 'molecular Aharonov–Bohm effect' in a different context, to describe how degeneracies in electron energy levels affect the spectrum of nuclear vibrations. He explains two options, both leading to the same vibration spectrum. The first option is to continue the electronic states round degeneracies (in the space of nuclear coordinates) in the manner described in this paper, thus causing the electronic wavefunctions to be multi-valued, with a compensating multi-valuedness in the nuclear states, which must be incorporated into their boundary conditions. The alternative is to enforce single-valuedness on the electronic (and hence also the nuclear) states, and this introduces a vector potential into the Schrödinger equation for nuclear motion. In general one may expect such effects whenever an isolated system is considered as being divided into two interacting parts, each slaved to a different aspect of the other (in the molecular case, electron states are slaved to nuclear coordinates, and nuclear states are slaved to the electronic states and wavefunctions). The systems considered in this paper might be regarded as a special case, in which the coupling is with 'the rest of the Universe' (including us as observers). The only role of the rest of the Universe is to provide a Hamiltonian with slowly-varying parameters, thus forcing the system to evolve adiabatically with phase continuation governed by the time-dependent Schrödinger equation.

## 6. Discussion

A system slowly transported round a circuit will return in its original state; this is the content of the adiabatic theorem. Moreover its internal clocks will register the passage of time; this can be regarded as the meaning of the dynamical phase factor. The remarkable and rather mysterious result of this paper is that in addition the system records its history in a deeply geometrical way, whose natural formulation (9) and (10) involves phase functions hidden in parameter-space regions which the system has not visited.

The total phase of the transported state (5) is dominated by the dynamical part, because $T \to \infty$ in the adiabatic limit, and it might be thought that this must overwhelm the geometrical phase $\gamma_n$ and make its physical effects difficult to detect. This objection can be met by observing that the strengths of non-adiabatic transitions are exponentially small in $T$ if $\hat{H}$ changes smoothly (Hwang & Pechukas 1977), so that essentially adiabatic evolution can occur even when the dynamical phase is only a few times greater than $2\pi$.

As we saw in § 3, degeneracies in the spectrum of $\hat{H}(\boldsymbol{R})$ are the singularities of the vector $\boldsymbol{V}(\boldsymbol{R})$ (equation (10)) in parameter space, and so have an important effect on the geometrical phase factor. This is reminiscent of the part played by singularities of an analytic function, but the analogy is imperfect: if $\gamma(\mathrm{C})$ were completely singularity-determined, $\boldsymbol{V}(\boldsymbol{R})$ would be the sum of the 'magnetic fields' of 'monopoles' situated at the degeneracies (cf. (17)) and so would have zero curl, which is not the case (zero curl, unlike zero divergence, is not a property which is invariant under deformations of $\boldsymbol{R}$ space, and in the general case the sources of $\boldsymbol{V}$ are not just monopoles but also 'currents' distributed continuously in parameter space). A closer analogy is with wavefront dislocation lines, which are phase singularities of complex wavefunctions in three-dimensional position space (Nye & Berry 1974; Nye 1981; Berry 1981), that dominate the geometry of wavefronts without completely determining them.

In view of the emphasis on degeneracies as organizing centres for phase changes, it is worth remarking that close approach of energy levels is not a necessary condition for the existence of nontrivial geometrical phase factors. Indeed, our examples have shown that $\gamma(\mathrm{C})$ can be non-zero even if C involves isospectral deformations of $\hat{H}(\boldsymbol{R})$ (in the Aharonov–Bohm illustration, the levels $E_n$ do not depend on $\boldsymbol{R}$ at all).

The results obtained here are not restricted to quantum mechanics, but apply more generally, to the phase of eigenvectors of any Hermitian matrices under a natural continuation in parameter space. Therefore they have implications throughout wave physics. For example, the electromagnetic field of a single mode travelling along an optical fibre will change sign if the cross section of the fibre is slowly altered so that its path (in the space of shapes) surrounds a shape for which the spectrum of the Helmholtz equation is degenerate (such as one of the diabolical triangles discovered by Berry & Wilkinson 1984).

I thank Dr J. H. Hannay, Dr E. J. Heller and Dr B. R. Pollard for several suggestions, and Professor Barry Simon for showing me, before publication, his paper which comments on this one. This work was not supported by any military agency.

## References

Aharonov, Y. & Bohm, D. 1959 *Phys. Rev.* **115**, 485–491.
Aharonov, Y. & Susskind, L. 1967 *Phys. Rev.* **158**, 1237–1238.
Arnold, V. I. 1978 *Mathematical methods of classical dynamics*. New York: Springer.
Berry, M. V. 1980 *Eur. J. Phys.* **1**, 240–244.
Berry, M. V. 1981 Singularities in waves and rays. In *Les Houches Lecture Notes for session* XXXV (ed. R. Balian, M. Kléman & J-P. Poirier), pp. 453–543. Amsterdam: North-Holland.
Berry, M. V. & Wilkinson, M. 1984 *Proc. R. Soc. Lond.* A **392**, 15–43.
Chambers, R. G. 1960 *Phys. Rev. Lett* **5**, 3–5.
Ehrenburg, W. & Siday, R. E. 1949 *Proc. phys. Soc.* B **62**, 8–21.
Herzberg, G. & Longuet-Higgins, H. C. 1963 *Discuss. Faraday Soc.* **35**, 77–82.
Hwang, J.-T. & Pechukas, P. 1977 *J. chem. Phys.* **67**, 4640–4653.
Longuet-Higgins, H. C. 1975 *Proc. R. Soc. Lond.* A **344**, 147–156.
Mead, C. A. 1979 *J. chem. Phys.* **70**, 2276–2283.
Mead, C. A. 1980*a* *Chem. Phys.* **49**, 23–32.
Mead, C. A. 1980*b* *Chem. Phys.* **49**, 33–38.
Mead, C. A. & Truhlar, D. G. 1979 *J. chem. Phys.* **70**, 2284–2296.
Messiah, A. 1962 *Quantum mechanics*, vol. 2. Amsterdam: North-Holland.
Morandi, G. & Menossi, E. 1984 *Nuovo Cim.* B (Submitted.)
Nye, J. F. 1981 *Proc. R. Soc. Lond.* A **378**, 219–239.
Nye, J. F. & Berry, M. V. 1974 *Proc. R. Soc. Lond.* A **336**, 165–190.
Rauch, H., Wilfing, A., Bauspiess, W. & Bonse, U. 1978 *Z. Phys.* B **29**, 281–284.
Rauch, H., Zeilinger, A., Badurek, G., Wilfing, A., Bauspiess, W. & Bonse, U. 1975 *Physics Lett.* A **54**, 425–427.
Schulman, L. S. 1981 Techniques and Applications of Path Integration. New York: John Wiley.
Silverman, M. P. 1980 *Eur. J. Phys.* **1**, 116–123.
Simon, B. 1983 *Phys. Rev. Lett.* (In the press.)
Stone, A. J. 1976 *Proc. R. Soc. Lond.* A **351**, 141–150.
Von Neumann, J. & Wigner, E. P. 1929 *Phys. Z.* **30**, 467–470.
Werner, S. A., Colella, R., Overhauser, A. W. & Eagen, C. F. 1975 *Phys. Rev. Lett.* **35**, 1053–1055.
Wu, T. T. & Yang, C. N. 1975 *Phys. Rev.* D **12**, 3845–3857.

## Appendix A

To show that $\gamma(C)$ is independent of the surface spanning C, it is necessary to prove that $\boldsymbol{V}(\boldsymbol{R})$ (equation (10)) has zero divergence. This can be accomplished by expressing $\boldsymbol{V}$ in terms of the vector Hermitian operator $\hat{\boldsymbol{B}}$ defined by

$$\hat{\boldsymbol{B}} \equiv -\mathrm{i} \sum_n |\nabla n\rangle \langle n|. \tag{A 1}$$

From (8), the off-diagonal elements of $\hat{\boldsymbol{B}}$ are

$$\langle n| \hat{\boldsymbol{B}} |m\rangle = -\mathrm{i} \langle m| \nabla H |n\rangle / (E_n - E_m), \quad m \neq n. \tag{A 2}$$

Thus (10) becomes

$$\boldsymbol{V} = \mathrm{Im}\,\langle n|\hat{\boldsymbol{B}} \times \hat{\boldsymbol{B}}|\, n\rangle. \tag{A 3}$$

Now we can calculate the divergence:

$$\nabla \cdot \boldsymbol{V} = \mathrm{Im}\,\{\langle \nabla n| \cdot \hat{\boldsymbol{B}} \times \hat{\boldsymbol{B}}\,|n\rangle + \langle n|\, \boldsymbol{B} \times \boldsymbol{B} \cdot |\nabla n\rangle + \langle n|\, \nabla \cdot (\hat{\boldsymbol{B}} \times \hat{\boldsymbol{B}})\,|n\rangle\}, \tag{A 4}$$

Use of a consequence of (A 1), namely

$$|\nabla n\rangle = \mathrm{i}\hat{\boldsymbol{B}}\,|n\rangle \tag{A 5}$$

gives

$$\nabla \cdot \boldsymbol{V} = n(-\hat{\boldsymbol{B}} \cdot \hat{\boldsymbol{B}} \times \hat{\boldsymbol{B}} + \hat{\boldsymbol{B}} \times \hat{\boldsymbol{B}} \cdot \hat{\boldsymbol{B}})\,|n\rangle + \mathrm{Im}\,\langle n|\, (\nabla \times \hat{\boldsymbol{B}} \cdot \hat{\boldsymbol{B}} - \hat{\boldsymbol{B}} \cdot \nabla \times \boldsymbol{B}\,|n\rangle. \tag{A 6}$$

For the curl of $\boldsymbol{B}$, (A 1) and (A 5) give

$$\nabla \times \hat{\boldsymbol{B}} = +\mathrm{i} \sum_n |\nabla n\rangle \times \langle \nabla n| = \mathrm{i} \sum_n \hat{\boldsymbol{B}}\,|n\rangle \times \langle n|\hat{\boldsymbol{B}} = \mathrm{i}\hat{\boldsymbol{B}} \times \hat{\boldsymbol{B}}\rangle, \tag{A 7}$$

whence $\nabla \cdot \boldsymbol{V}$ vanishes by the dot-cross rule for triple products.

This result is valid everywhere except at the 'monopole' singularities arising from degeneracies.